Optimal Smoothing of Gaussian Periodic Data

GARRETT BIRKHOFF & ARTHUR S. PRIVER

1. Introduction. A basic problem of the “calculus of observations” [18, Ch. XI] consists in *fitting* discrete data by as “smooth” a function as possible, *interpolated* through the data points. This problem can often be solved by standard interpolation formulas, by using broken-line, spline [1], [8], or other piecewise polynomial functions on a physical, semilog, or loglog scale.

A more general problem arises when the data points have statistical scatter, and when they involve errors of observation (human or instrumental) and/or computation (*e.g.*, roundoff). When the data points are widely spaced, and the scatter and probable errors are small, interpolation may still be the most practical solution. However, when the scatter or probable errors are appreciable and many data points are available, it is usually better to use a *smoothing* scheme based on least squares *approximation* and/or averaging to reduce the probable error.

Many “smoothing” schemes have been proposed in the literature. Clearly, the effectiveness of such smoothing schemes depends on the expected nature or “prior distribution” of (i) the sample function being measured and (ii) the probable measurement errors of the observations. Given this prior distribution, one can define a smoothing scheme to be *optimal* when it minimizes the expected mean square error; and one can then try to determine this optimal scheme *theoretically* by Bayesian techniques, using conditional probabilities.

In the present paper, we solve the resulting optimality problem in what is probably the simplest case: that of *periodic data* measured at equally spaced intervals, and having a homogeneous *normal* (Gaussian) prior distribution, under the assumption that measurement errors are also normally distributed and independent; see §2. The *optimal smoothing* scheme is then given by an easily computed convolution kernel, as we show in §5. (In some design problems, the “optimal” solution may combine closeness of fit with smoothness, for esthetic reasons; see [1, §13].)

Unfortunately, the practical value of the preceding theoretical result is very limited, because *the prior distribution is almost never known in practice*, even for periodic data measured at equally spaced intervals. Even when the prior distribution may be presumed to satisfy the conditions described above, one does not know what the relevant variances are.

In §6, we therefore return to the *practical* problem of smoothing periodic data measured at equally spaced intervals, *when the prior distribution is unknown*.

We describe briefly a computer program, written and tested by one of us [14], which has proved very helpful in solving this practical problem. Finally, in §7, we discuss in detail its practical applicability to data associated with a periodic Poisson process.

It is our conclusion that *interactive* computer programs, preferably with *graphic* capabilities like that described in [14], have more potential for practical data smoothing than the static *a priori* schemes usually proposed ([4], [10], [15], [18], [22]). Even when these contain a free parameter, such as the degree of an algebraic or trigonometric polynomial to be fitted by least squares, or the number of joints (“knots”) in a spline, they tend to be arbitrary unless the Bayesian aspects of the problem are explicitly recognized. And even when these Bayesian aspects are recognized (as in [8] and [19]), human judgment is usually required to evaluate the plausibility of an apparently optimal prior distribution. This judgment can, we think, be exercised most efficiently using an interactive computer program having graphic capabilities.

2. Smoothing schemes. Let a function $u(\theta)$ be measured at $N = 2n$ equally spaced points $\theta_j = 2\pi j/N = \pi j/n$ ($j = 0, 1, \dots, N - 1$), and let u_i be the value *measured* at $\theta = \theta_i$. Many schemes $\mathbf{S}$ have been proposed for smoothing the “measured” values u_i (data points), so as to give “smoothed” values

$$(1) \quad u_k^* = S_k[u_1, \dots, u_N] = S_k[\mathbf{u}]$$

having smaller probable errors; see for example [5, pp. 314–17], [10], [11, pp. 130–40], and [17, pp. 254–8].

Two of the most widely recommended kinds of smoothing schemes for periodic data are *convolution smoothing* and *truncation* of the Fourier series expansion of the measured values. These may be described in general terms as follows.

Convolution smoothing. By a “convolution smoothing” scheme for smoothing periodic data is commonly meant a scheme of the form

$$(2) \quad u_k^* = \sum_i w_{i-k} u_i = \sum_l w_l u_{k+l}, \quad w_l \geq 0, \quad \sum w_l = 1.$$

Such convolution smoothing schemes are also called “filtering” schemes (see Hamming [5] and Wiener [22]). See also §5, paragraph two below. Clearly, convolution smoothing is a *linear* smoothing scheme in the sense that

$$(3) \quad S_k[\mathbf{u} + \mathbf{v}] = S_k[\mathbf{u}] + S_k[\mathbf{v}], \quad k = 0, 1, \dots, N - 1.$$

Most commonly used convolution schemes are *symmetric*, in the sense that $w_{-l} = w_l$; thus this is true of the scheme described by Lanczos [9, pp. 316–20] with $S_k[\mathbf{u}] = u_k - (3/35)\delta^4 u_k$, and of the simpler scheme $u_k^* = (u_{k-1} + 2u_k + u_{k+1})/4$. However, Hamming [5] has proposed replacing u_k by $(u_k + u_{k+1} + u_{k+2})/3$. In general, convolution schemes tend to reduce the magnitude of the coefficients of the high-frequency terms in the Fourier series for $u(\theta)$.

Truncating Fourier series. More directly, one can simply cut off or *truncate* the high-frequency terms in the Fourier series expansion for the measured

values $u(\theta_i)$; see Quade and Collatz [15] and Lanczos [9, 10, 11]. Here one first interpolates

$$(4) \quad a_0/2 + \sum_{k=1}^{n-1} (a_k \cos k\theta + b_k \sin k\theta) + (a_n/2) \cos n\theta$$

through the set of observed values u_i by using the coefficients

$$(4') \quad \begin{aligned} a_k &= (2/N) \sum_{i=0}^{N-1} u_i \cos k\theta_i, & k &= 0, 1, \dots, n, \\ b_k &= (2/N) \sum_{i=0}^{N-1} u_i \sin k\theta_i, & k &= 1, \dots, n-1; \end{aligned}$$

see Jackson [6, pp. 109–15]. It is easily verified that the series replaces each term $a_l \cos l\theta$ or $b_l \sin l\theta$ with $l > n$ in the “true” Fourier series of a function $u(\theta)$ by the “equi-interpolatory” function $a_k = a_{l-rn}$ giving the least non-negative $k \equiv l \pmod{n}$. Clearly also, the substituted term $a_k \cos k\theta$ is “smoother” than $a_l \cos l\theta$ in the sense that the integral of the square of its m -th derivative is less for every $m > 1$. Although it does not follow that the net effect is to reduce $\int |g^{(m)}(\theta)|^2 d\theta$ for all $m > 1$, the general tendency is to “smooth” nevertheless, especially for m large (even for $m = 3$).

Such truncation schemes are also linear if N is fixed. However, the choice of the cutoff or “truncation” point in (4') is usually made *a posteriori*, and in general smoothing schemes with a variable truncation point are *nonlinear*.

Thus Lanczos suggests [9, p. 337] that all terms should be ignored as soon as the coefficients “have no tendency to become either much larger or much smaller than a certain average size.” Likewise Ralston [17, p. 235] suggests that the coefficients be ignored as soon as they no longer cause a “significant decrease” in the estimated variance of the errors of observation.

3. A Bayesian model. Somewhat as in [19], we shall use Bayesian techniques to analyze the relative merits of smoothing schemes such as those described in §2. Accordingly we shall assume that the “true” function is a sample from a class of random periodic functions given (as in [7, p. 187]) by

$$(5) \quad f(\theta) = \alpha_0/2 + \sum_{k=0}^{n-1} (\alpha_k \cos k\theta + \beta_k \sin k\theta) + (\alpha_n/2) \cos n\theta.$$

We shall assume that each measured value u_i is the sum $u_i = f(\theta_i) + e_i$ of the “true” value $f(\theta_i)$ given by setting $\theta = \theta_i = 2\pi j/N$ in (5), and a “measurement error” (error of observation) e_i . For mathematical simplicity, we shall make the following assumptions on the prior distribution of the function (5) being measured and on the measurement errors $u_k - f(\theta_k) = e_k$.

Assumption 1. The coefficients α_k and β_k in (5) are *normal* random variables having *independent* distributions with mean zero and known variances η_k .

(In most applications, the η_k are not known precisely, but may be assumed to tend fairly rapidly to zero as $k \rightarrow \infty$.)

Assumption 2. The measurement errors e_k have independent Gaussian distributions with mean zero and variance $\bar{e^2} = \epsilon$. That is, in the notation of [20, p. 57], we assume that the independent errors of measurement e_i have normal distributions of the form $N(0, \epsilon)$.

Under these assumptions, we shall derive theoretically the smoothing scheme giving the *best fit* to the *sample function* $f(\theta)$ being measured, in the sense that it *minimizes the mean square error*. Because of our assumptions, the resulting, *optimal smoothing scheme* $\mathbf{u} \rightarrow \mathbf{S}[\mathbf{u}]$ turns out to be a *linear convolution* scheme as will be shown.

To this end, we shall next see how the errors e_i affect the coefficients a_k, b_k of the Fourier series (4)–(4') which interpolates to the u_k . Since $u_i = f(\theta_i) + e_i$, we see by comparing (5) with (4)–(4') that the coefficients of the Fourier series for the random errors e_i satisfy $a_k = \alpha_k + \alpha'_k, b_k = \beta_k + \beta'_k$. We can compute these coefficients using Assumption 2 as follows.

Lemma 1. *Under Assumption 2, the measurement errors are given by $e_i = e(\theta_i)$, where the periodic function $e(\theta)$ is defined by the random Fourier series*

$$(6) \quad e(\theta) = \alpha'_0/2 + \sum_{k=1}^{n-1} (\alpha'_k \cos k\theta + \beta'_k \sin k\theta) + (\alpha'_n/2) \cos n\theta,$$

whose coefficients α'_k and β'_k are independent normal random variables with distributions $N(0, 2\epsilon/N)$ for $k = 1, \dots, n-1$ and $N(0, 4\epsilon/N)$ for α'_0, α'_n .

Proof. Let $\mathbf{w}^T = [\alpha'^T \beta'^T]$. We may write (4)–(4') in the form $\mathbf{w} = M\mathbf{e}$, where M is nearly orthogonal. The e_i are independent Gaussian random variables, and since α'_k and β'_k are linear combinations of the e_i , they are consequently Gaussian random variables. Now consider the means

$$(7) \quad \bar{\mathbf{w}} = E(\mathbf{w}) = ME(\mathbf{e}) = \mathbf{0}.$$

Thus all the error distributions have mean zero.

Next consider the covariance matrix

$$(8) \quad E(\mathbf{w} - \bar{\mathbf{w}})(\mathbf{w} - \bar{\mathbf{w}})^T = E(M\mathbf{e}\mathbf{e}^T M^T) = ME(\mathbf{e}\mathbf{e}^T)M^T.$$

Since the errors e_i are independent, $E(\mathbf{e}\mathbf{e}^T) = \epsilon I$. Now the matrix ϵMM^T is readily seen by orthogonality to be a diagonal matrix with all elements $2\epsilon/N$, except the first and the $(n+1)$ st, which are $4\epsilon/N$. Hence the covariances are all zero; since the distributions are all Gaussian, the coefficients are therefore independent random variables. Q.E.D.

4. Bayesian analysis. The likelihood function for a random variable is [13, p. 10–12] proportional to a *conditional probability* density function which expresses the likelihood of occurrence of any value of a random variable, given

the “prior” information that a certain measurement or measurements (observations) were made subject to some error distribution. Bayes’ Theorem [13, p. 10–10] asserts that the conditional posterior distribution is the product of the prior distribution and the likelihood function, divided by the integral of this product. Since we are given the prior distributions of the coefficients of (1), we can apply Lemma 1 to determine the probable effect of smoothing.

To make this application, we need one additional result, which expresses the (suitably normalized) product of two Gaussian distributions as a Gaussian distribution [13, p. 16–11].

Lemma 2. *Let $h(\omega) = CN(a, \nu)N(b, \eta)$, where C is a normalizing constant defined to make $\int h(\omega)d\omega = 1$. Then $h(\omega)$ is a normal distribution with mean $(a\eta + b\nu)/(\eta + \nu)$ and variance $\nu\eta/(\nu + \eta)$. The mean of $h(\omega)$ is simply a weighted average of the means a and b , the weights being the reciprocals of the corresponding variances. The reciprocal of the variance of $h(\omega)$ is the sum of the reciprocals of the variances ν and η .*

We now apply Bayes’ Theorem and Lemmas 1–2 to obtain the conditional posterior distributions for each coefficient in (5). For each coefficient, we know the (normal) prior distribution, the likelihood function for the N measurements u_i , and the (normal) prior distributions of the errors of measurement. Substituting from these into Bayes’ Theorem, we obtain the following new result.

Theorem 1. *Let Assumptions 1–2 hold, and let $\nu = 2\epsilon/N$. Let $\mu_k = 1/(1 + M_k)$ where*

$$(9a) \quad M_k = \begin{cases} N\eta_k/2\epsilon = \eta_k/\nu & \text{for } k = 1, \dots, n-1, \\ N\eta_k/4\epsilon = \eta_k/2\nu & \text{for } k = 0, n \end{cases},$$

and let

$$(9b) \quad p_k = 1 - \mu_k = 1 - (1/[1 + M_k]) = M_k/(1 + M_k).$$

Then, after one set of measurements u_i , the coefficients α_k and β_k in (5) have Gaussian (conditional) posterior distributions $N(\bar{\alpha}_k, V_k)$ and $N(\bar{\beta}_k, V_k)$, where

$$(9c) \quad \bar{\alpha}_k = p_k a_k, \quad \bar{\beta}_k = p_k b_k, \quad V_k = \mu_k \nu_k.$$

The proof is a straightforward application of Bayes’ Theorem, using Lemma 2 to calculate the posterior distributions from the prior distributions given by Assumption 1 and Assumption 2 (as reformulated in Lemma 1).

5. Optimal smoothing. Theorem 1 gives the (normal) posterior conditional probability distribution of the Fourier coefficients α_k, β_k of the sample function (5) after one set of values u_i has been measured; these values are also normally distributed [20, p. 99]. We shall now show that the *optimal smoothing scheme* consists in taking the average values $\bar{\alpha}_k$ and $\bar{\beta}_k$ of these posterior coefficients from (9c) to generate “smoothed values”, so that

$$(10) \quad u_i^* = S_i[\mathbf{u}] = \bar{\alpha}_0/2 + \sum_{k=1}^{n-1} (\bar{\alpha}_k \cos k\theta_i + \bar{\beta}_k \sin k\theta_i) + (\bar{\alpha}_n/2) \cos n\theta_i,$$

where $\bar{\alpha}_k = p_k a_k$ and $\bar{\beta}_k = p_k b_k$ are given by (9a)–(9c).

Note that, in this case, optimal smoothing is equivalent by (10) to a simple *linear filter*, which diminishes the amplitude of Fourier components progressively as the frequency increases, without changing their phase (except possibly by 180° , which surprises us).

Here by “optimal” or “best” we mean that the expected cost of the resulting error is least, for a wide class of *loss functions* $\lambda(f, w)$ as discussed (for example) by Ferguson [3, p. 1]. We shall consider the class Λ of loss functions $\lambda(f - w) = \lambda(q)$ satisfying the three conditions:

- (i) $\lambda(0) = 0$,
- (ii) $\lambda(q) = \lambda(-q)$ (symmetric),
- (iii) If $q \geq q' \geq 0$, then $\lambda(q) \geq \lambda(q')$.

(Note that choosing $\lambda(q) = q^2$ corresponds to minimizing the mean square error.) Then the following theorem, attributed to Sherman (see Deutsch [2, pp. 16–17] for the proof), states that if the distribution of f is symmetric and if either the distribution of f or the loss function λ is convex then the best estimate of f is the mean $\bar{f}$ of the distribution. Under these conditions, (10) is a direct consequence of this result.

Theorem 2. *Let x be a random variable whose conditional distribution $F(x | \mathbf{u})$, given a set of observations $\mathbf{u} = (u_1, \dots, u_n)$, is symmetric about its mean $\bar{x} = E(x | \mathbf{u})$. Let $\lambda \in \Lambda$ be any monotone symmetric loss function. If either the conditional distribution $F(x | \mathbf{u})$ is convex for $x \leq \bar{x}$ or the loss function λ is convex, then the conditional expectation $E(x | \mathbf{u})$ minimizes the risk function*

$$R(x) = E(\lambda(x - w) | \mathbf{u})$$

where $w = \phi(u_1, \dots, u_n)$ is a known function.

In the particular case of a loss function $\lambda(x - w) = |x - w|^m$, $m \geq 1$, even when the distribution of x is not symmetric about $\bar{x}$, $E(|x - w|^2)$ is minimized by setting $w = \bar{x}$ (Yule and Kendall [21, p. 155]), and $E(|x - w|)$ is minimized when w is the median of x (Lindgren [12, p. 63]) (which is $\bar{x}$ for any normal or other symmetric distribution).

From Theorems 1 and 2, we obtain the following result as an immediate consequence.

Corollary. *Under Assumptions 1–2 with given η_k and ϵ , the optimal smoothing scheme is obtained by multiplying the Fourier coefficients a_k, b_k of the measured values by the factors p_k defined in (9a)–(9b); it is linear.*

The limiting special case $\eta_0 = \infty$, $p_0 = 1$ is very important for applications; it corresponds to complete absence of information about the zero-point of measurement (the mean value of the sample function).

6. Convolution smoothing. We shall now show that the optimal smoothing scheme given by (10) is equivalent to convolution of the “measured” values u_i with a fixed discrete kernel γ_{i-k} . Indeed, it is obvious from rotational symmetry that the “smoothing” scheme defined by (9c) is invariant under any rotation $\theta \rightarrow \theta + 2\pi j/N$; it is also linear (see Greville [4, p. 157]). On the other hand we have the following

Theorem 3. *Any linear smoothing scheme for measured $u(\theta_k) = u_k$ which is invariant under rotation through $2\pi/N$ is of the “convolution” form*

$$(11) \quad u_k^* = \sum_{l=0}^{N-1} \gamma_l u_{k+l}, \quad (k, l \bmod N).$$

It is invariant also under reflection if and only if $\gamma_l = \gamma_{N-l}$ (symmetry condition).

Proof. First define the smoothing operators at the individual data points. Let $\mathbf{i}_0 = (1, 0, \dots, 0)$ and $\mathbf{i}_j$ be the $(j+1)$ st unit vector. Let $S[\mathbf{i}_0] = (\gamma_0, \dots, \gamma_{N-1})$ be the smoothing operator at the zeroth point. Then by rotational invariance

$$S[\mathbf{i}_i] = \sum \gamma_l \mathbf{i}_{i+l} = \sum \gamma_{k-i} \mathbf{i}_k = (\gamma_{-i}, \gamma_{-i+1}, \dots, \gamma_{N-i-1}),$$

with all subscripts mod N . Hence for any $\mathbf{u} = \sum u_k \mathbf{i}_k$,

$$u_i^* = S[\mathbf{i}_i](\mathbf{u}) = (\sum \gamma_{k-i} \mathbf{i}_k) \cdot (\sum u_k \mathbf{i}_k) = \sum_k \gamma_{k-i} u_k = \sum_l \gamma_l u_{i+l}. \quad \text{Q.E.D.}$$

To compute the weights γ_l in (11), one simply expands the function $u_0 = 1$, $u_j = 0$ ($j \neq 0$) in a discrete truncated cosine (Fourier) series. From (4)–(4'), we get $a_k = 2/N$ ($k = 0, \dots, n$) and $b_k = 0$ ($k = 1, \dots, n-1$). Hence, by (9),

$$(12) \quad \gamma_i = \frac{2}{N} \sum_{k=1}^{n-1} \left\{ \frac{\cos j\theta_k}{1 + (\nu/\eta_k)} \right\} + \frac{1}{N} \left\{ \frac{1}{1 + (2\nu/\eta_0)} + \frac{\cos j\theta_n}{1 + (2\nu/\eta_n)} \right\}.$$

Letting $c_{ik} = \cos j\theta_k = \cos(2\pi jk/N)$, and setting $V_k = \mu_k \eta_k$ from (9c), we can alternatively write

$$(13) \quad \gamma_i = \frac{2}{\nu N} \sum_{k=1}^{n-1} V_k c_{ik} + \frac{1}{N} \left\{ \frac{V_0}{2\nu} + \frac{V_n c_{in}}{2\nu} \right\}.$$

To simplify (13), we let $\sum''$ mean “sum, taking half of the first and last summands,” getting

$$(13') \quad \gamma_i = \frac{2}{N} \sum_{k=0}^n {}'' p_k c_{ik}, \quad j = 0, \dots, N-1.$$

We see from (13) that the weights γ_i are symmetric (invariant under the transformation $\theta \rightarrow -\theta$), i.e., $\gamma_i = \gamma_{-i}$. There are thus $n+1$ distinct weights γ_i to calculate, from the $n+1$ given prior variances η_k . Typical numerical

values of the γ_i are tabulated in Table 1. (Table 1, and several other results of this paper, are taken from Arthur Priver's Doctoral Thesis [14].)

$N = 16$

j	p						
	1	2	3	4	5	6	7
0	0.994427	0.970375	0.869201	0.672783	0.501810	0.398047	0.336157
1	0.002639	0.019426	0.092418	0.212920	0.275524	0.280015	0.265989
2	-0.000601	-0.007719	-0.037633	-0.046151	0.014912	0.080494	0.120488
3	0.000858	0.005741	0.019632	-0.000192	-0.034436	-0.025292	0.004513
4	-0.000605	-0.005067	-0.016534	-0.007501	-0.013774	-0.034828	-0.039783
5	0.000719	0.004768	0.014713	0.008234	0.004442	-0.010600	-0.031427
6	-0.000605	-0.004619	-0.014023	-0.007121	0.001946	0.003847	-0.007102
7	0.000688	0.004549	0.013590	0.006798	0.001456	0.005497	0.010878
8	-0.000606	-0.004527	-0.013520	-0.006751	-0.001948	0.003689	0.016735

$N = 24$

j	p						
	1	2	3	4	5	6	7
0	0.994629	0.958943	0.785183	0.524722	0.366488	0.284374	0.237265
1	0.002423	0.025937	0.139600	0.256294	0.264222	0.237554	0.211269
2	-0.000399	-0.008689	-0.038021	0.010648	0.095560	0.135843	0.147016
3	0.000640	0.005894	0.012652	-0.020089	-0.002264	0.041861	0.073232
4	-0.000403	-0.004936	-0.011666	-0.011340	-0.026807	-0.012449	0.014280
5	0.000498	0.004500	0.009605	0.001839	-0.016331	-0.027578	-0.018931
6	-0.000403	-0.004266	-0.008927	-0.000827	-0.003623	-0.020366	-0.028265
7	0.000460	0.004129	0.008302	0.001976	0.002607	-0.007459	-0.022506
8	-0.000403	-0.004043	-0.008041	-0.001583	0.002558	0.001502	-0.011280
8	0.000445	0.003990	0.007799	0.001481	0.001357	0.004451	-0.001178
10	-0.000404	-0.003956	-0.007699	-0.001497	-0.000059	0.003280	0.005184
11	0.000440	0.003939	0.007610	0.001469	-0.000192	0.001121	0.008111
12	-0.000404	-0.003933	-0.007606	-0.001459	-0.000542	0.000107	0.008872

TABLE 1. Weights γ_i for $\eta_k = k^{-p}$, $\eta_0 = \infty$, $\epsilon = .01$

Discussion. A casual inspection of Table 1 reveals some interesting facts. First the sum of the "weighting" coefficients γ_i of the convolution kernel is unity, as expected. However, this is only true when $\eta_0 = \infty$.

More surprising is the fact that, for all the optimal smoothing schemes associated with the variances $\eta_k = k^{-p}$, *some weights are negative*; this may be because a *small* $\epsilon = .01$ was assumed. However, we do not have simple conditions on the η_k and ϵ which are necessary to make all $\gamma_k \geq 0$. A desire to understand the facts better led us to pose some interesting new questions.

Converse questions. Up to now, we have taken the prior distribution of the function $f(\theta)$ as given, and have determined the choice of convolution kernel which would minimize the effect of measurement errors (for the class of models considered). We now ask the converse question: given a (symmetric) convolution

kernel $\gamma = (\gamma_0, \dots, \gamma_{N-1})$, what are the distributions of functions and measurement errors to which it is well-suited?

Mathematically, the problem is to invert (13) so as to express the p_k and hence the η_k in Assumption 1 (or better, the η_k/ϵ) in terms of the γ_i and ν . The inversion is given by

Theorem 4. *Given a symmetric convolution kernel $\gamma = (\gamma_0, \dots, \gamma_{N-1})$, define*

$$(14) \quad p_0 = \frac{1}{N} \sum_{i=0}^{N-1} \gamma_i = \frac{2}{N} \sum_{i=0}^n {}'' \gamma_i,$$

and, more generally,

$$(15) \quad p_k = \frac{1}{N} \sum_{i=0}^{N-1} c_{ki} \gamma_i = \frac{2}{N} \sum_{i=0}^n {}'' c_{ki} \gamma_i, \quad k = 0, \dots, n.$$

(We do not assume here that $\gamma_i \geq 0$ or that $\sum \gamma_i = 1$.) Then, if $0 \leq p_i \leq 1$ for all j , the given convolution kernel γ is optimal for the prior distribution of §2 with mean zero and variances

$$(16) \quad \eta_k = \begin{cases} \nu M_k = \nu p_k / (1 - p_k) = \nu p_k / \mu_k & \text{for } k = 1, \dots, N-1, \\ 2\nu M_k = 2\nu p_k / (1 - p_k) = 2\nu p_k / \mu_k & \text{for } k = 0, N, \end{cases}$$

with the p_k in (9c) given by (14)–(15).

The proof follows directly from (13) and the orthogonality property of the discrete cosine functions.

Corollary. *Under Assumptions 1–2, the sum of the coefficients of the optimal smoothing kernels (“weights”) is*

$$(17) \quad \sum_i \gamma_i = 1/[1 + (4\epsilon/N\eta_0)] = 1/[1 + (2\eta/\eta_0)].$$

Thus, for $\epsilon > 0$, the convolution scheme of Theorem 3 satisfies the convolution condition $\sum \gamma_i = 1$ if and only if $\eta_0 = \infty$.

7. Practical applications. In practice, very little is usually known about the “prior distribution” of the sample function being measured. Hence the formulas derived above are primarily of theoretical interest.

In practice, since the optimality of a given smoothing scheme depends on the assumed prior distribution, which is unknown, one must therefore consider the selection of the η_k and ϵ as a (Bayesian) *hypothesis* to be tested. Such testing can be very efficient on a computer, and one of us [14] has developed an interactive computer graphics system for making such tests. Periodic statistical data were smoothed with the help of the formulas of §§2–6 using the following technique. First, we adopted as figures of merit (cf. [18, p. 304])

$$(18) \quad R = \frac{1}{N} \sum_{i=1}^N (u_i - u_i^*)^2 \quad \text{and} \quad S = \sum_{i=1}^{N-2} (\Delta^3 u_i)^2,$$

since these are standard measures of goodness of fit and smoothness, respectively. We then tried out a range of hypotheses concerning the *a priori* Bayesian distribution. *It proved possible in all cases to improve the data-smoothing procedures used by the authors of the papers on the subject* which we found, with respect to both smoothness and closeness of fit! For a detailed discussion, see [14].

Poisson prior distributions. Some of the data which we studied came from what was presumably a non-Gaussian Poisson distribution. Nevertheless, the procedure for *a posteriori* data smoothing described above proved very successful. We will conclude this paper with a few theoretical comments on this case.

For a Poisson process, the probability that n events will occur in a given time interval for which the expected number is a , is given by

$$(19) \quad p_n = e^{-a} a^n / (n!).$$

An easy computation shows that the *variance* $\overline{(n - a)^2} = \sum p_n (n - a)^2$ of a Poisson distribution is always equal to its *mean*.

For example, consider the diurnal variation of the distribution of thunderstorms in each hour of the day (most in the afternoon). The number n_i of days on which such a thunderstorm occurs in the j -th hour may be assumed to have a Poisson distribution with mean $\overline{n_i}$; suppose that we try to fit an observed (periodic) function n_i by a discrete Fourier series, smoothing the data so as to get a better approximation to the "true" (Bayesian) average a_i as a convolution, getting $n_i^* = \sum \gamma_k n_{i+k}$.

Then, since the *true* expected variance satisfies

$$\sum \overline{(n_i - a_i)^2} = \sum a_i,$$

we must view with suspicion any "smoothing" process whose deviation satisfies

$$\sum (n_i^* - n_i)^2 < \sum n_i.$$

Any significantly better closeness of fit must be considered as either fortuitous or spurious!

REFERENCES

1. G. BIRKHOFF & CARL DE BOOR, Piecewise polynomial interpolation and approximation in *Approximation of Functions*, ed. H. Garabedian, Elsevier, Amsterdam, 1965, 164–190.
2. R. DEUTSCH, *Estimation Theory*, Prentice-Hall, 1964.
3. T. S. FERGUSON, *Mathematical Statistics—A Decision Theoretic Approach*, Academic Press, New York, 1967.
4. T. N. E. GREVILLE, *On stability of linear smoothing formulas*, SIAM J. Numer. Anal.—B3 (1966), 157–170.
5. R. W. HAMMING, *Numerical Methods for Scientists and Engineers*, McGraw-Hill, New York, 1962.
6. D. JACKSON, *The Theory of Approximation*, Amer. Math. Soc., New York, 1930.
7. J.-P. KAHANE, *Some Random Series of Functions*, D. C. Heath, 1968.

8. G. S. KIMELDORF & G. WAHBA, *Bayesian estimation on stochastic processes and smoothing by splines*, Ann. Math. Stat. **41** (1970), 495–502.
9. C. LANCZOS, *Applied Analysis*, Prentice Hall, 1956.
10. C. LANCZOS, *Evaluation of noisy data*, SIAM J. Numer. Anal.—B1 (1964), 76–85.
11. C. LANCZOS, *Discourse on Fourier Series*, Hafner Publishing Co., New York, 1966.
12. B. W. LINDGREN, *Statistical Theory*, Macmillan Co., New York, 1962.
13. J. W. PRATT, H. RAIFFA, & R. SCHLAIFER, *Introduction to Statistical Decision Theory*, McGraw-Hill, New York, 1965.
14. A. S. PRIVER, *Data Smoothing in Interactive Computer Graphics*, Doctoral thesis, Division of Engineering and Applied Physics, Harvard University, Cambridge, Mass., 1970.
15. W. QUADE & L. COLLATZ, *Zur interpolationstheorie der reelen periodischen funktionen*, S-B Preuss. Akad. der Wiss, 1938, 383–429.
16. H. RAIFFA & R. SCHLAIFER, *Applied Statistical Decision Theory*, Division of Research, Graduate School of Business Administration, Harvard University, Boston, 1961.
17. A. RALSTON, *A First Course in Numerical Analysis*, McGraw-Hill, New York, 1965.
18. SIR EDMUND WHITTAKER & G. ROBINSON, *The Calculus of Observations*, 4th ed., Blackie and Son Ltd., London, 1944.
19. P. WHITTLE, *On the smoothing of probability densities*, J. Roy. Stat. Soc. **20** (1958), 334–343.
20. S. C. WILKS, *Mathematical Statistics*, Princeton Univ. Press, Princeton, 1943.
21. G. UDNEY YULE & M. G. KENDALL, *An Introduction to the Theory of Statistics*, 12th ed., J. B. Lippincott Co., London, 1940.
22. C. H. REINSCH, *Numerische Math.*, **10** (1967), 177–183.

Harvard University

Date Communicated: MARCH 8, 1971